



THE UNIVERSITY OF  
CHICAGO

PEDIATRIC CANCER  
DATA COMMONS

# Transforming the Way Researchers Share Data

LESSONS FROM THE PEDIATRIC CANCER DATA COMMONS

Slides available at  
<http://sam.am/CAC2-2020>

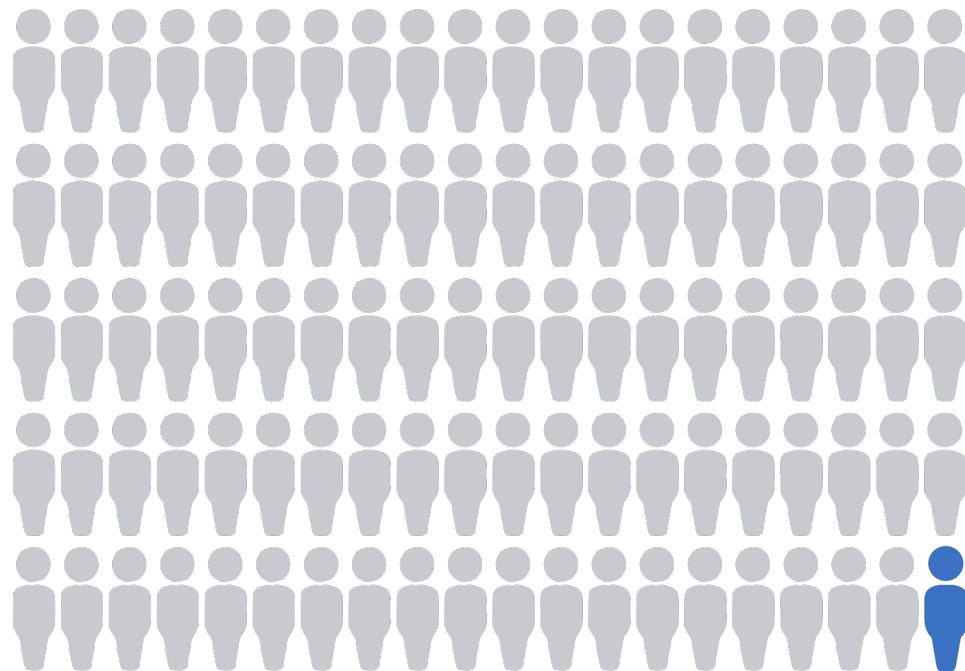
# Learning points

- **Data standards:** Data standards are a fundamental building block in harnessing the potential of big data.
- **Pediatric-specific challenges:** Pediatric cancer presents unique challenges and is lagging behind, in large part because of the lack of pediatric-specific data standards – particularly clinical data standards.
- **Consensus:** Defining and maintaining data standards require deliberation, debate, and consensus from across the pediatric cancer spectrum and medical specialties.
- **Ecosystems and collaboration:** Harnessing the potential of big data requires an ecosystem of federated systems that can work together to advance the cause.

# A Perfect Storm: Rare diseases, siloed data, and manual processes



# Pediatric cancer is a rare disease



18 million new cases  
of cancer worldwide  
every year

← 224,000 (**1%**)  
are children

Sources: [Globocan 2018](#) (adult), [Ward 2019](#) (pediatric)

# Ultimately, this is what we're trying to fix



| October                   |                                                                        |                                                                                        |                                                               |                                                   |                                   |                           |
|---------------------------|------------------------------------------------------------------------|----------------------------------------------------------------------------------------|---------------------------------------------------------------|---------------------------------------------------|-----------------------------------|---------------------------|
| William                   |                                                                        |                                                                                        |                                                               |                                                   |                                   |                           |
| Sun                       | Mon                                                                    | Tue                                                                                    | Wed                                                           | Thu                                               | Fri                               | Sat                       |
|                           | <b>1</b> Bactrim<br>Cytoxin 20mg                                       | <b>2</b> Bactrim<br>Cytoxin 20mg                                                       | <b>3</b> cycle 2<br>DTR 9:30am<br>Avastin IV<br>Cytoxin 20mg  | <b>4</b><br>DTR 9:30am<br>Cytoxin IV<br>Zometa IV | <b>5</b><br>Cytoxin 20mg          | <b>6</b><br>Cytoxin 20mg  |
| <b>7</b><br>Cytoxin 20mg  | <b>8</b> Bactrim<br>Cytoxin 20mg<br>Labs and blood<br>pressure at home | <b>9</b> Bactrim<br>Cytoxin 20mg                                                       | <b>10</b><br>Cytoxin 20mg                                     | <b>11</b><br>Cytoxin 20mg                         | <b>12</b><br>Cytoxin 20mg         | <b>13</b><br>Cytoxin 20mg |
| <b>14</b><br>Cytoxin 20mg | <b>15</b> Bactrim<br>Cytoxin 20mg                                      | <b>16</b> Bactrim<br>Cytoxin 20mg                                                      | <b>17</b><br>DTR 9:30am<br>Avastin IV<br>Cytoxin 20mg         | <b>18</b><br>Cytoxin 20mg                         | <b>19</b><br>Cytoxin 20mg         | <b>20</b><br>Cytoxin 20mg |
| <b>21</b><br>Cytoxin 20mg | <b>22</b> Bactrim<br>Cytoxin 20mg<br>SSKI                              | <b>23</b> Bactrim<br>Cytoxin 20mg<br>MIBG, x-ray knees<br>Bone marrow biopsies<br>SSKI | <b>24</b><br>Cytoxin 20mg<br>MIBG<br>CT scan<br>SSKI          | <b>25</b><br>Cytoxin 20mg<br>SSKI                 | <b>26</b><br>Cytoxin 20mg<br>SSKI | <b>27</b><br>Cytoxin 20mg |
| <b>28</b><br>Cytoxin 20mg | <b>29</b> Bactrim<br>Cytoxin 20mg                                      | <b>30</b> Bactrim<br>Cytoxin 20mg                                                      | <b>31</b> Cycle 3<br>DTR 9:30am<br>Avastin IV<br>Cytoxin 20mg | <b>1</b><br>DTR 9:30am<br>Cytoxin IV<br>Zometa IV |                                   |                           |

# Current state: Lack of standardization

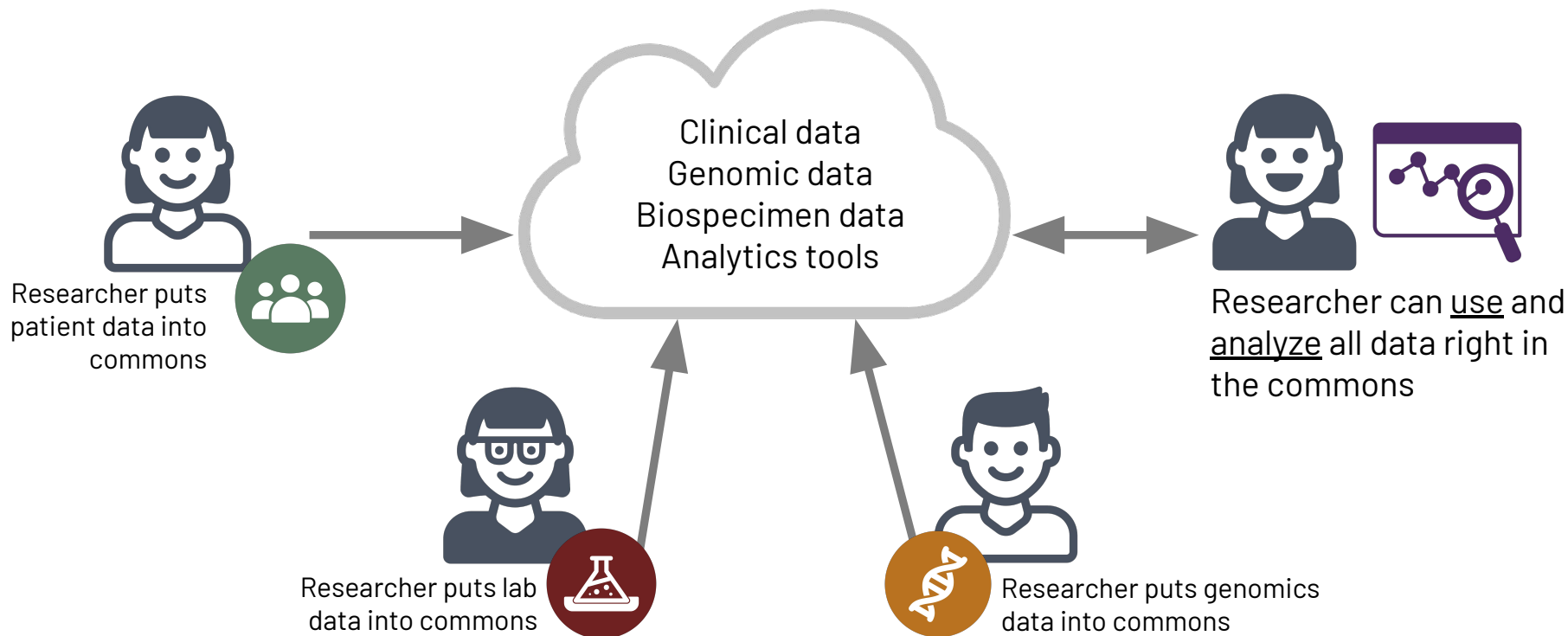
Ways that age is expressed in the Gene Expression Omnibus (GEO)

|                          |                        |
|--------------------------|------------------------|
| <i>age</i>               | <i>age [y]</i>         |
| <i>Age</i>               | <i>age [year]</i>      |
| <i>AGE</i>               | <i>age [years]</i>     |
| <i>`Age</i>              | <i>age in years</i>    |
| <i>age (after birth)</i> | <i>age of patient</i>  |
| <i>age (in years)</i>    | <i>Age of patient</i>  |
| <i>age (y)</i>           | <i>age of subjects</i> |
| <i>age (year)</i>        | <i>age(years)</i>      |
| <i>age (years)</i>       | <i>Age(years)</i>      |
| <i>Age (years)</i>       | <i>Age(yrs.)</i>       |
| <i>Age (Years)</i>       | <i>Age, year</i>       |
| <i>age (yr)</i>          | <i>age, years</i>      |
| <i>age (yr-old)</i>      | <i>age, yrs</i>        |
| <i>age (yrs)</i>         | <i>age.year</i>        |
| <i>Age (yrs)</i>         | <i>age_years</i>       |

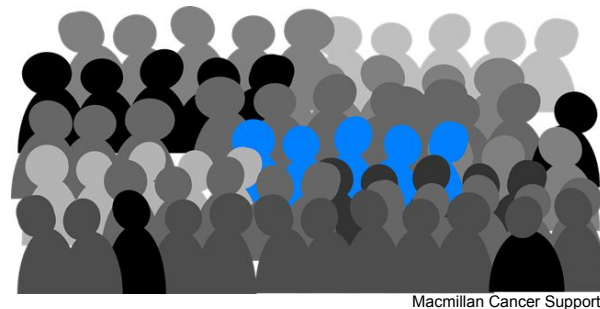
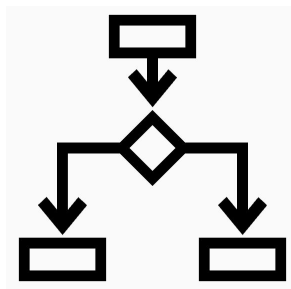
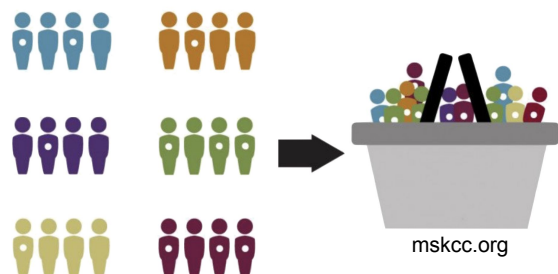
  
Gene Expression Omnibus

Adapted from Mark A. Musen, M.D., Ph.D.

# What is a data commons?



# What kinds of research can a commons enable?



- Correlating **biomarkers** with clinical outcomes across trials
- Understanding impact of **dose modifications** across trials
- Performing patterns of failure analyses
- Examining **toxicity** prognosticators

- Validating consensus **staging** definitions across trials
- Validating **prognostic scoring** systems
- Enhancing **risk stratification**

- **Prognosis** of rare subgroups
- **Age-related differences** in therapeutic response
- **Survivorship** studies
- **Disparities** analyses

# Essential elements for building a commons



## CONSORTIUM

building trust between groups



## SCIENTIFIC GOALS

why build it? what data to include?



## DATA GOVERNANCE

publication policy, approving data use



## DATA DICTIONARY

everyone speaking the same language



## DATA TRANSFORMATION AND AGGREGATION

statisticians and data scientists

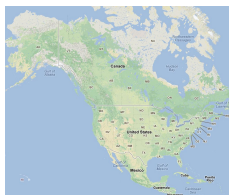


## FUNDING

building and sustaining the commons

# Our first commons (2015): Neuroblastoma

## INTERNATIONAL NEUROBLASTOMA RISK GROUP (INRG), 2004



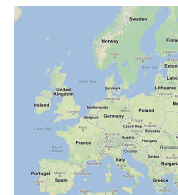
## Children's Oncology Group (COG)



Germany



Japan



SIOPEN

| Group   | Number |
|---------|--------|
| COG     | 4235   |
| Germany | 1938   |
| Japan   | 470    |
| SIOPEN  | 936    |
| Total   | 8800   |

The good news: 8800 patients

[illegible]

## The bad news: data in Excel

# Neuroblastoma data commons - Cohort discovery

## Cohort Search

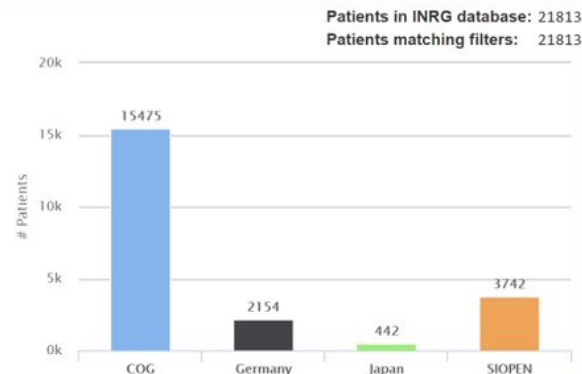
Search: New Save Search

Add a filter: - Please select - Clear Filters

- Please select -
- Primary Tumor-Pelvis
- Primary Tumor-Thorax
- Race
- Revised INPC Prognostic Group/ Shimada Diagnostic Category
- Site of Relapse
- Time from Dx to Death or Last Contact
- Time from Dx to Event or Last Contact
- Year of Diagnosis
- Asgharzadeh GEO data - **Note! External Data**
- GEO Data - **Note! External Data**
- GWAS Data - **Note! External Data**
- Imaging Data - **Note! External Data**
- Nationwide Tissue Bank - **Note! External Data**
- Nucleic Acid Data - **Note! External Data**
- TARGET Data - **Note! External Data**

External data

## Results



Publicly available at <http://inrgdb.org/>



# Imaging data

National Cancer Institute  
at the National Institutes of Health | www.cancer.gov

**NBIA National Biomedical Imaging Archive**

Home | Search Images | Manage Data Basket | Tools | Admin | User's Guide | Logout | bturner

**Simple Search**

Add all found to

Results Per Page: 20

Collection(s): ANBL1221

Image Modality(ies): ALL ☒ ANY Modality

Return cases that include:

☐ CR ☐ CT ☐ MR ☐ NM

| Collection ID | Subject ID | Matched Studies | Total Studies | Matched Series | Total Series |
|---------------|------------|-----------------|---------------|----------------|--------------|
| ANBL1221      | PAURPL     | 5               | 5             | 25             | 25           |
| ANBL1221      | PAVCCI     | 6               | 6             | 64             | 64           |
| ANBL1221      | PAVCIM     | 6               |               |                |              |
| ANBL1221      | PAVFKI     | 4               |               |                |              |
| ANBL1221      | PAVBHD     | 2               |               |                |              |

National Cancer Institute  
at the National Institutes of Health | www.cancer.gov

**NBIA National Biomedical Imaging Archive**

Home | Search Images | Manage Data Basket | Tools | Admin | User's Guide | Logout | bturner

**Search Results (Studies for Subject PAURPL)**

Add all found to

| Study Instance UID                               | Description                         | Date     | Baseline           | Add This Study To Basket |            |           |        |
|--------------------------------------------------|-------------------------------------|----------|--------------------|--------------------------|------------|-----------|--------|
| 1.2.410.10001.2.0.100.10000002.303.2015010204840 | MR PELVIS WITH AND WITHOUT CONTRAST |          |                    |                          |            |           |        |
| Series                                           | Description                         | Modality | Manufacturer       | Images                   | Thumbnails | Cine mode | DISCOM |
| ...021168.26                                     | COR STR                             | MR       | GE MEDICAL SYSTEMS | 1                        |            | NA        |        |
| ...021168.27                                     | FSE T1 COR                          | MR       | GE MEDICAL SYSTEMS | 22                       |            |           |        |
| ...021168.28                                     | FSE T1 AX                           | MR       | GE MEDICAL SYSTEMS | 24                       |            |           |        |
| ...021168.29                                     | FSE T2 AX F18                       | MR       | GE MEDICAL SYSTEMS | 24                       |            |           |        |
| ...021168.30                                     | Seq FSE T2                          | MR       | GE MEDICAL SYSTEMS | 42                       |            |           |        |
| ...021168.31                                     | FSE T1 AX + C                       | MR       | GE MEDICAL SYSTEMS | 24                       |            |           |        |
| ...021168.32                                     | FSE T1 SAG + C                      | MR       | GE MEDICAL SYSTEMS | 42                       |            |           |        |
| ...021168.33                                     | FSE T1 COR + C                      | MR       | GE MEDICAL SYSTEMS | 22                       |            |           |        |

| Study Instance UID                                | Description                | Date     | Baseline             | Add This Study To Basket |            |           |        |
|---------------------------------------------------|----------------------------|----------|----------------------|--------------------------|------------|-----------|--------|
| 1.2.410.10001.2.0.100.10000002.366.20150115100410 | MBG WHOLE BODY Y 123 1 DAY |          | Baseline + 13 day(s) |                          |            |           |        |
| Series                                            | Description                | Modality | Manufacturer         | Images                   | Thumbnails | Cine mode | DISCOM |
| ...650000006                                      | Adrenal SAVE_SCREEN        | NM       | SIEMENS NM           | 1                        |            | NA        |        |
| ...590000012                                      | SAVE_SCREEN                | NM       | SIEMENS NM           | 9                        |            |           |        |



# Visualization and analytic tools

## Visualization

### Survival analysis parameters:

Survival time limit (in years): 0 - 10



Event Free/ Overall Survival:

Overall Survival

Select a factor variable:

Ethnicity

Select a stratification variable:

- Please select -

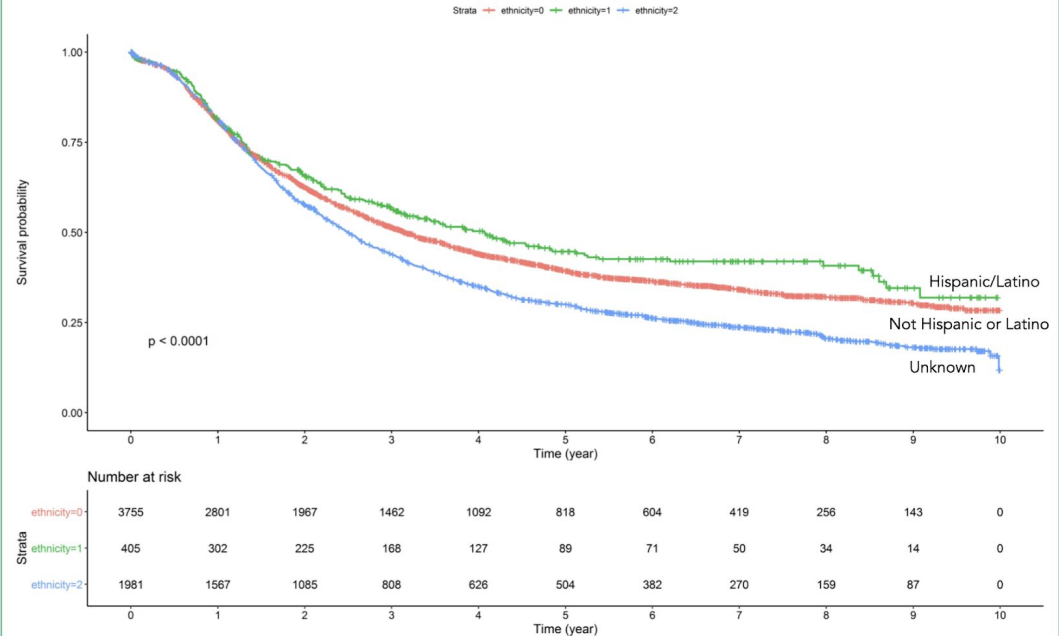
Calculate

### Cohort selection criteria:

INSSStage:

Stage 4

### Results



# Consensus-based decision making and data sharing

# International Soft Tissue Sarcoma Consortium (INSTRuCT)



Copenhagen  
May 2017

Chicago  
October 2017

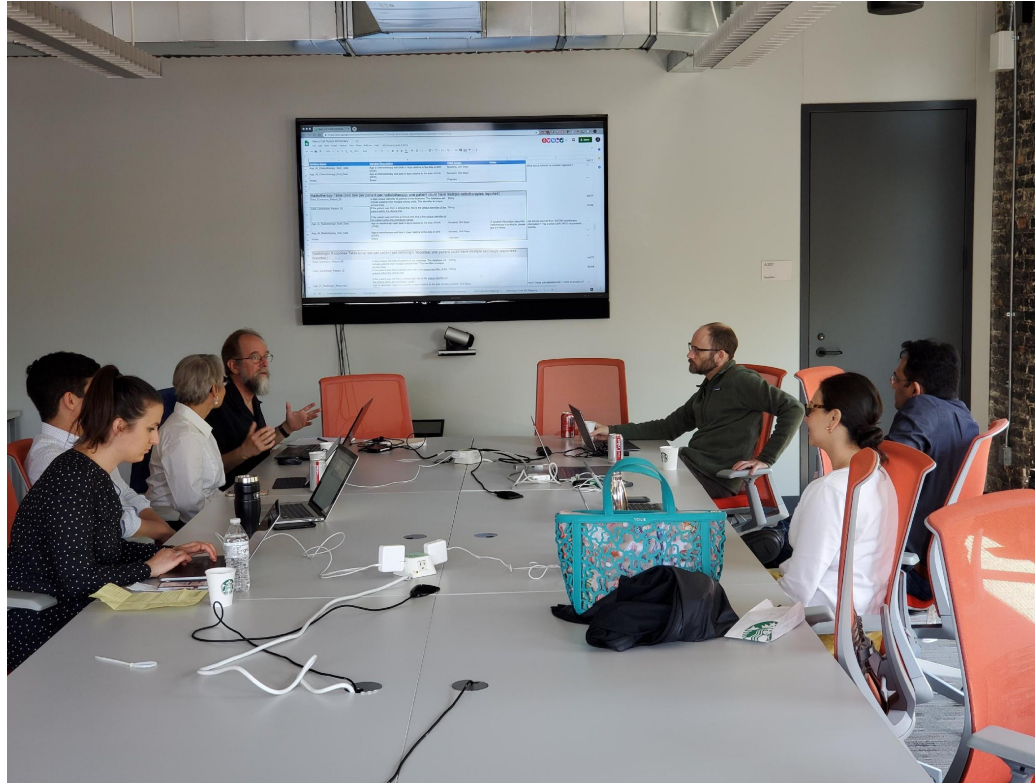
Amsterdam  
March 2018

Tübingen  
September 2018

Utrecht  
December 2018

Prague  
May 2019

# Data dictionary development

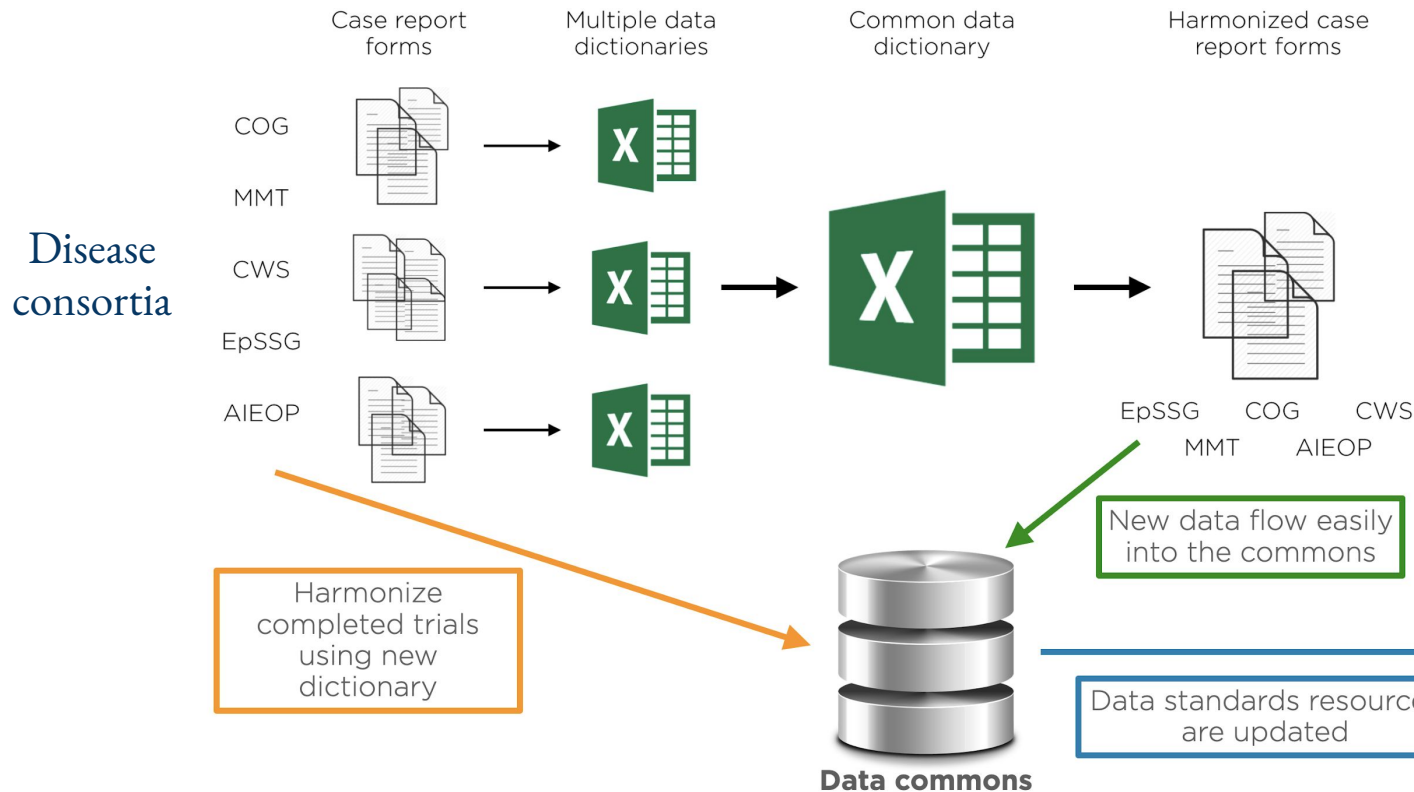




# Data dictionary development

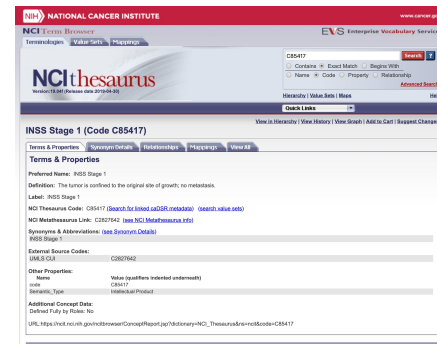


# Balloting a consensus data dictionary



# Data standards example

## Neuroblastoma stage



|              |          |                        |                                                            |
|--------------|----------|------------------------|------------------------------------------------------------|
| 1 = Stage 1  | Stage 1  | <a href="#">C85417</a> | The tumor is confined to the original site of growth; no   |
| 2 = Stage 2a | Stage 2a | <a href="#">C85418</a> | The tumor is unilateral and the resection is grossly incor |
| 3 = Stage 2b | Stage 2b | <a href="#">C85419</a> | The tumor is unilateral and the resection is complete or   |
| 4 = Stage 3  | Stage 3  | <a href="#">C85420</a> | The tumor extends across the midline and the regional l    |
| 5 = Stage 4  | Stage 4  | <a href="#">C85421</a> | Tumor spread to distant lymph nodes, bone marrow, b        |
| 6 = Stage 4s | Stage 4s | <a href="#">C85422</a> | Patients are less than one year old with localized primar  |
| 9 = Unknown  | Unknown  | <a href="#">C17998</a> | Not known, not observed, not recorded, or refused.         |



# Pediatric Cancer Data Commons Expansion

# Guiding principles governing the PCDC

## 1 OUR GOAL

is to lift barriers and connect researchers to data

## 2 STAKEHOLDER APPROVAL

for data release from any disease commons

## 3 CONTRIBUTOR APPROVAL

for data release from the original contributing group

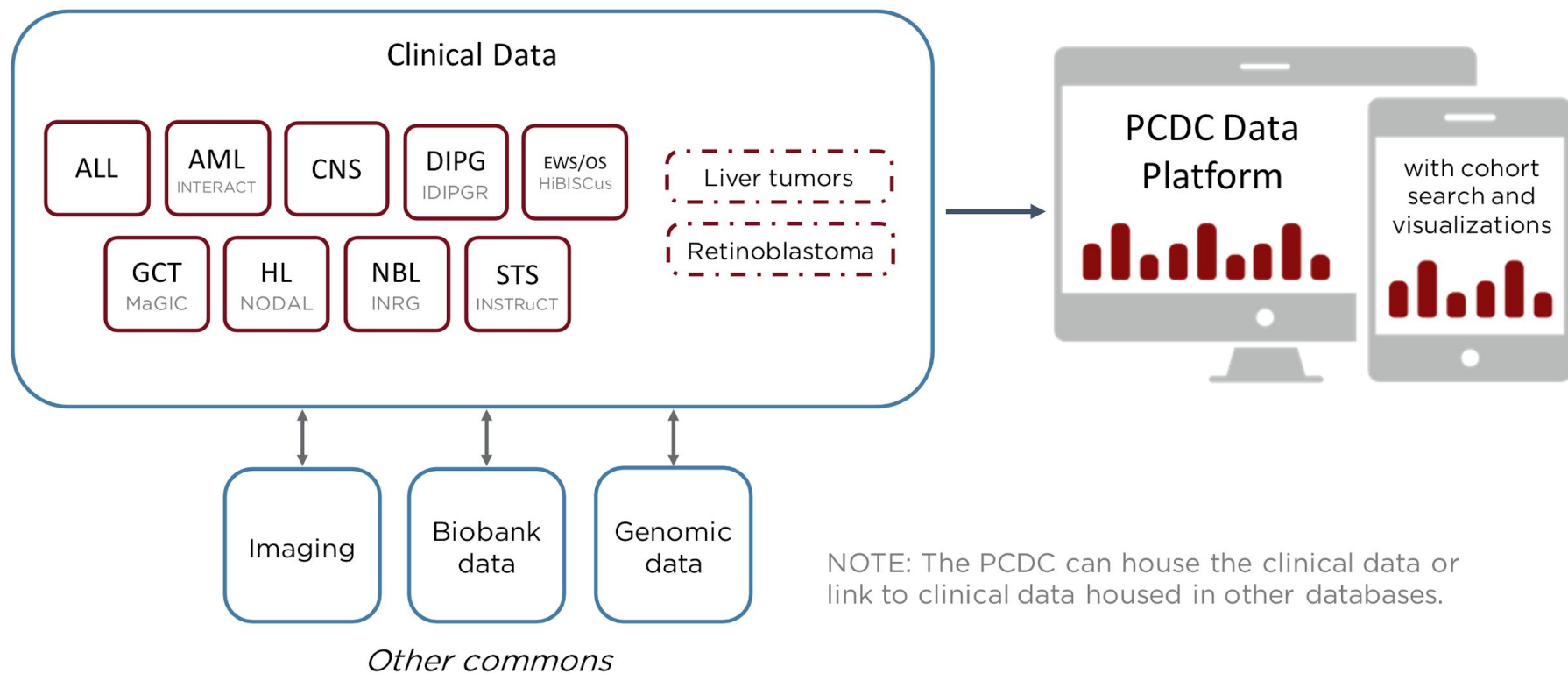
## 4 REPRESENTATION

on the Executive Committee from every disease group

## 5 RECOGNIZE REGIONAL DIFFERENCES

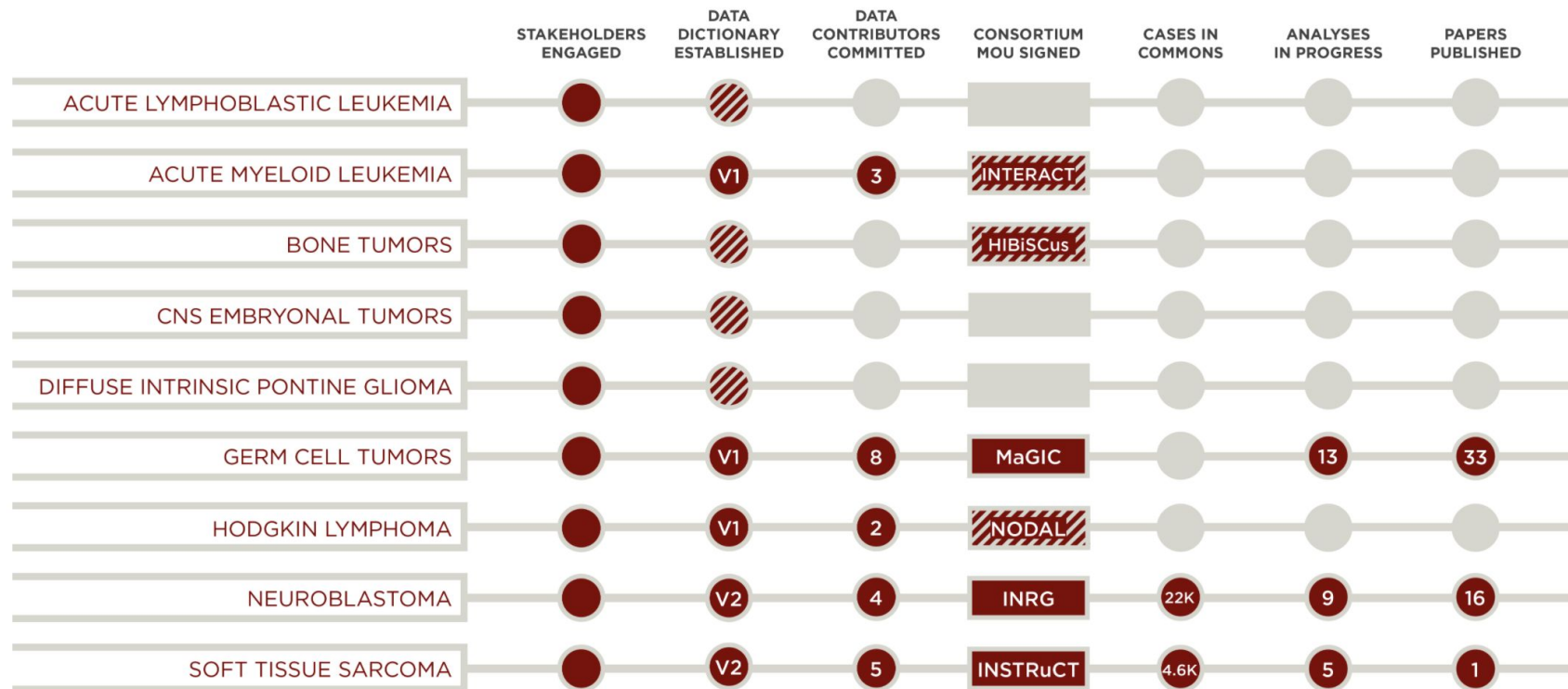
Privacy and data sharing laws vary by region

# An integrated pediatric cancer data commons



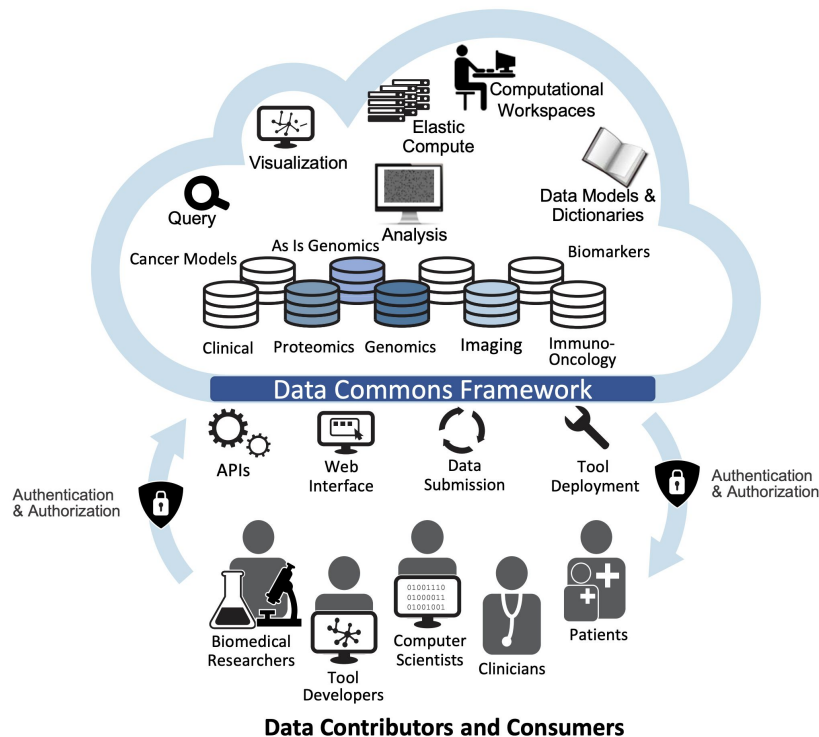
NOTE: The PCDC can house the clinical data or link to clinical data housed in other databases.

# PCDC progress to date



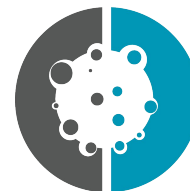
# Integration with national efforts

# NCI Cancer Research Data Commons ecosystem



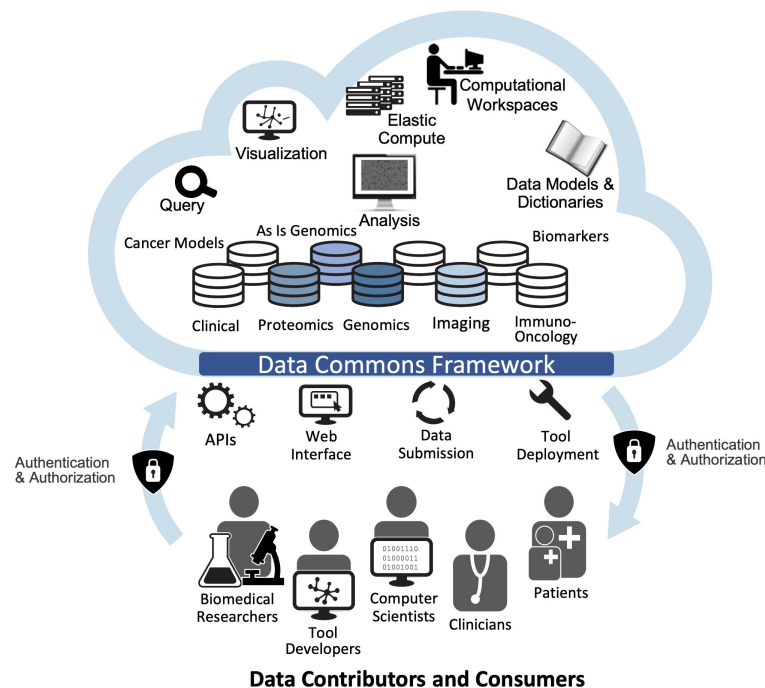
Credit: Allen Dearry

# Center for Cancer Data Harmonization (CCDH)



CENTER *for*  
CANCER DATA  
HARMONIZATION

- **Facilitate** retrospective and prospective semantic harmonization of data across nodes of the CRDC
- **Coordinate** the community to ensure implementation of standards that will facilitate interoperability of heterogeneous data types and CRDC resources
- **Find agreement** across the communities built around CRDC
  - match and extend data models
  - annotation, harmonization
  - quality assurance



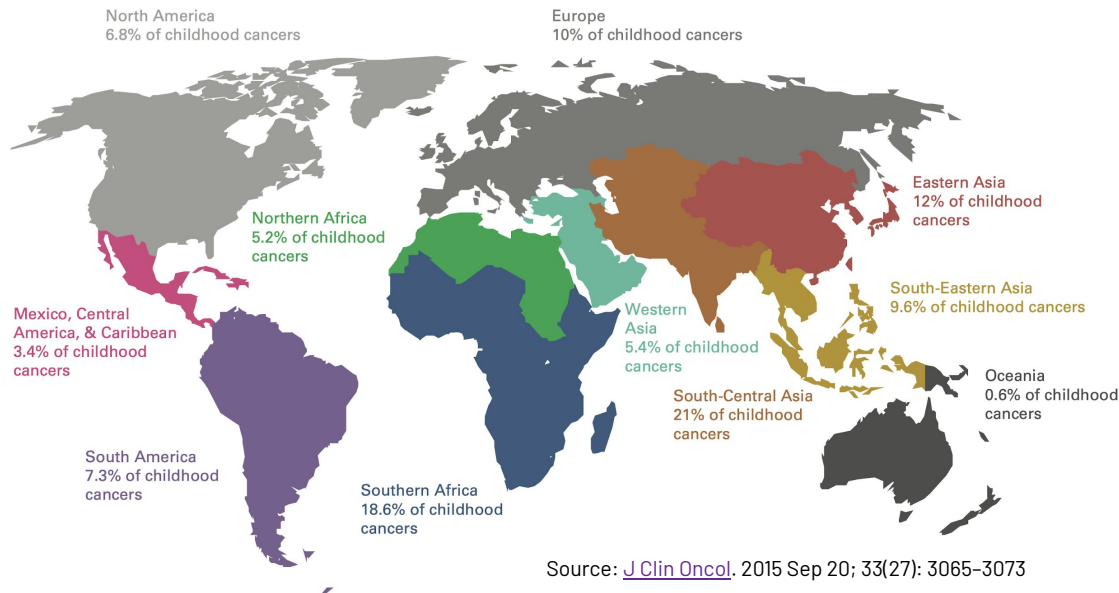
Credit: Allen Dearry



# Closing Global Gaps in Pediatric Cancer Survival

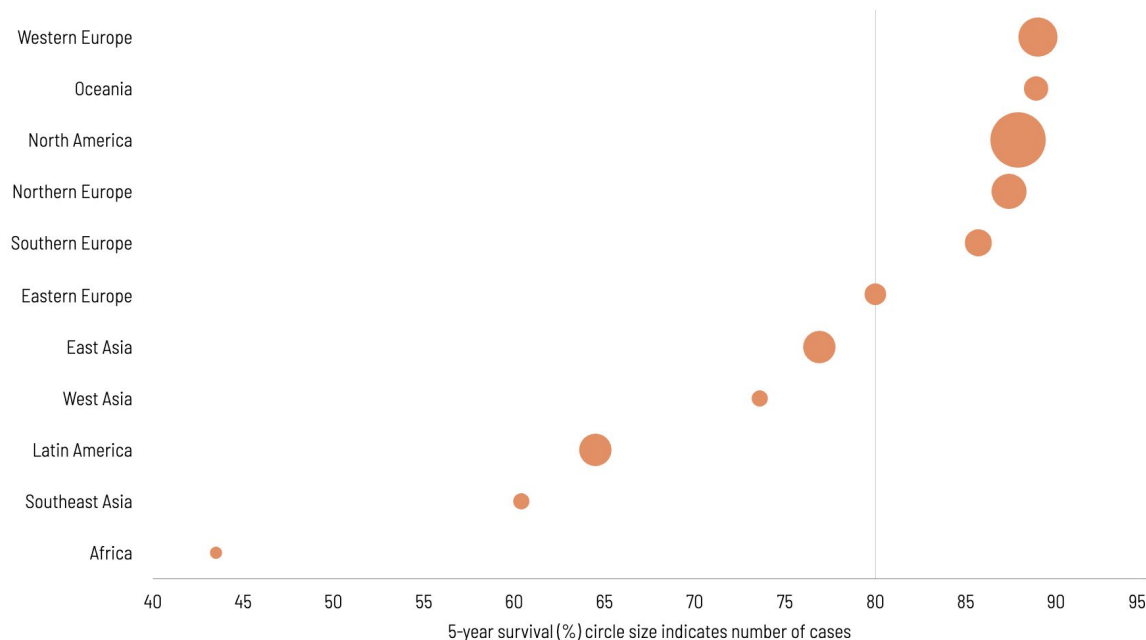
# Research to date does not benefit all children equally

- 80% of children with cancer diagnoses live in low- and middle-income countries
- Research to date does not capture diversity of the global population



# Survival rate disparity

Five-year age-standardized net survival (%) for lymphoid leukemia



<https://canceratlas.cancer.org/the-burden/cancer-in-children/?figure-chart=7WPF5EPc>

# Opportunities for the PCDC to close global gaps

- **Continue to promote and incorporate consensus data standards development globally**
- Recent impact:
  - INRG - recent publication
  - MaGIC - working with researchers

---

## Pediatric Blood & Cancer

---

ONCOLOGY: RESEARCH ARTICLE

The prognostic strength of serum LDH and serum ferritin in children with neuroblastoma: A report from the International Neuroblastoma Risk Group (INRG) project

Veronica Moroz, David Machin, Barbara Hero, Ruth Ladenstein, Frank Berthold, Paige Kao, Yaa Obeng, Andrew D.J. Pearson, Susan L. Cohn, Wendy B. London ✉

First published: 30 May 2020 | <https://doi.org/10.1002/pbc.28359>

### Conclusions

LDH and ferritin are strongly prognostic in NB, overall and within high-risk NB patients treated post-2009 with modern therapy. LDH and ferritin show promise for (a) identifying ultra-high-risk; (b) refining risk stratification; and (c) clinical utility in low-/middle-income countries. Routine collection of LDH and ferritin should be reinitiated for evolving NB risk stratification.

# Targeting Survivorship

# Survivorship

- A barrier to studying survivorship and long-term outcomes is access to longitudinal data
- PCDC is working with various stakeholders to create better ways to collect data on survivors
- Potential novel data collection methods include crowdsourcing
- A long-term outcomes / survivorship data commons is an essential component of our pediatric cancer strategy

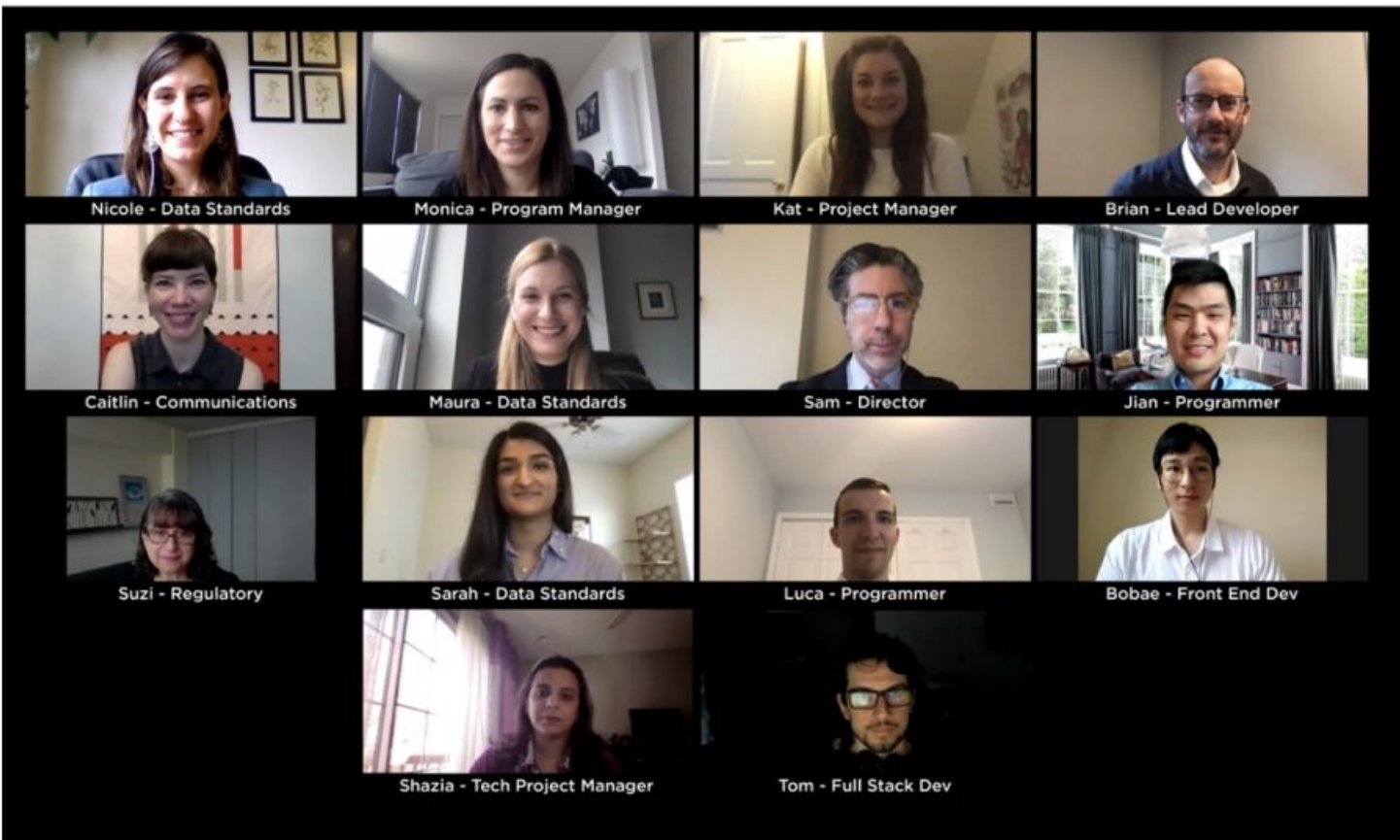
# Take home points

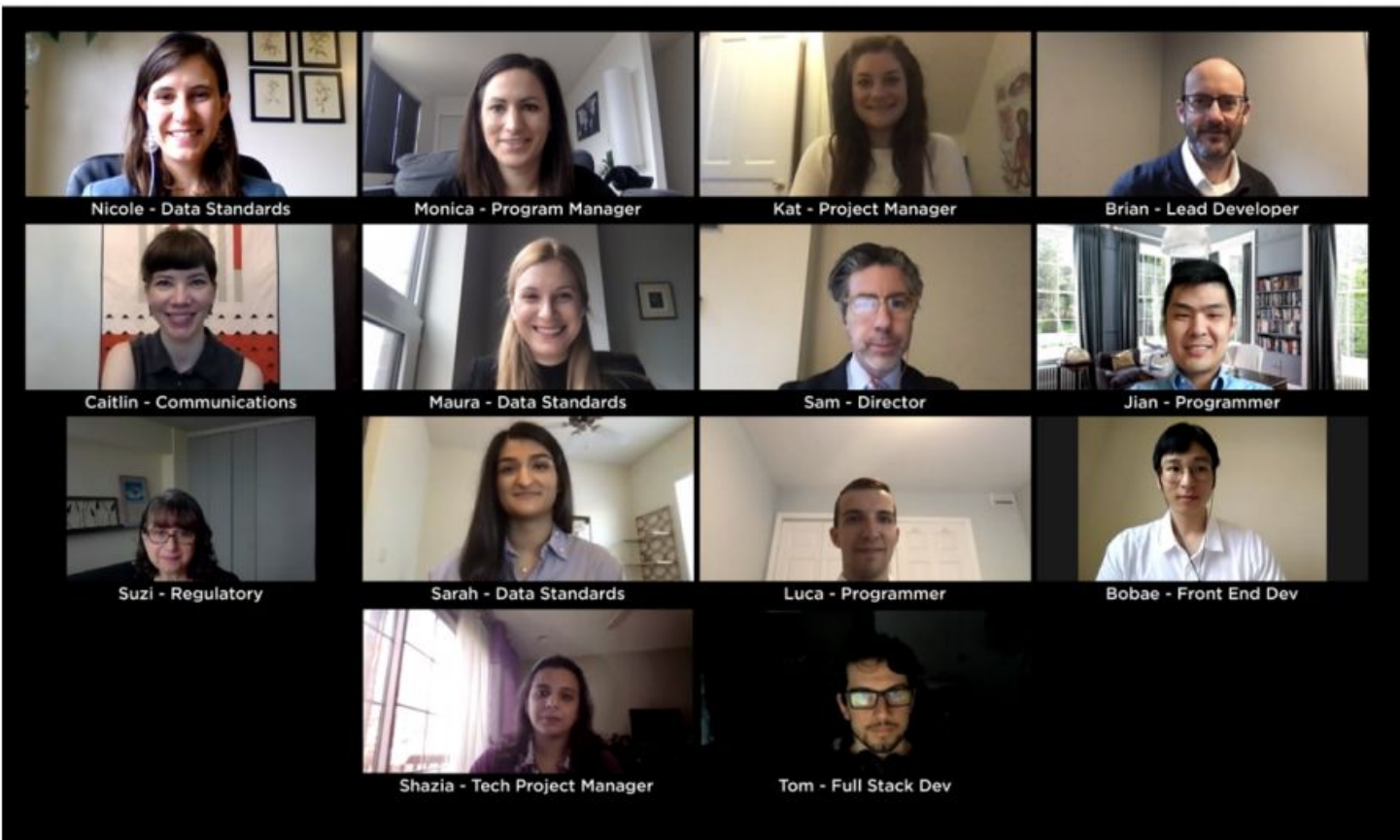
- Studying pediatric cancer requires **collaboration and sharing**
- Data sharing needs to be built on a foundation of **trust and consensus**
- **Connecting disparate data types** and sources enriches research
- **Consensus data standards** are critical for the success of national and international data ecosystems - allowing aggregation across trials and diseases
- **Early adoption of data standards** and consideration for the **lifecycle of the data** is critical to accelerating progress and discovery

# The Pediatric Cancer Data Commons team

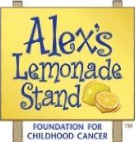








# We gratefully acknowledge our funders

|                                                                                                                                                                                                  |                                                                                                                                        |                                                                                                                                                                |                                                                                                                                                  |                                                                                                                                            |
|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----------------------------------------------------------------------------------------------------------------------------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------|--------------------------------------------------------------------------------------------------------------------------------------------|
|                                                                                                                 |  LEUKEMIA & LYMPHOMA SOCIETY®                         |  St. Baldrick's FOUNDATION<br>Conquer Childhood Cancers                       |  FOR CHILDHOOD CANCER RESEARCH<br><b>Rally!</b><br>FOUNDATION |  THE MATTHEW BITTKER FOUNDATION<br>It's What Matters    |
|  The Neuroblastoma Children's Cancer Society                                                                    |  WILLIAM GUY FORBECK RESEARCH FOUNDATION<br>wgfrt.org |  ANDREW McDONOUGH<br>B+ FOUNDATION®                                          |  <b>BRIGHTSIDE</b>                                            |  CANCER RESEARCH FOUNDATION                             |
|  <b>kick cancer</b>                                                                                             |  Children's Neuroblastoma Cancer Foundation           |  The Super Jake Foundation                                                   |  King Baudouin Foundation                                     |  KAT'S CREW                                             |
|  Alex's Lemonade Stand<br>FOUNDATION FOR CHILDHOOD CANCER                                                       |  CHILDREN'S RESEARCH FOUNDATION<br>"So they may live" |  LITTLE HEROES                                                               |  THE TED MULLIN FUND<br>Fighting To Beat Sarcoma              |  Seattle Children's<br>HOSPITAL • RESEARCH • FOUNDATION |
|  AT THE FOREFRONT OF <b>KIDS</b> MEDICINE®<br><b>UChicago Medicine</b><br>Comer Children's<br>Development Board |  The Brumfield Family                                 |  AT THE FOREFRONT<br><b>UChicago Medicine</b><br>Comprehensive Cancer Center | A gift made in memory of Payton O'Brien                                                                                                          |                                                                                                                                            |
|                                                                                                                                                                                                  |                                                                                                                                        |                                                                                                                                                                |  NIH <b>NATIONAL CANCER INSTITUTE</b>                         |                                                                                                                                            |

